

**Express Mail Label No. EL960712630US**

Date of Deposit: 10.08.2003

**APPLICATION FOR LETTERS PATENT  
OF THE UNITED STATES**

**NAME OF INVENTOR(S):**

Dorin Comaniciu  
1108 Bradley Court  
Princeton, NJ 08540  
Citizen: UNITED STATES OF AMERICA

**TITLE OF INVENTION:**

Density Estimation-Based Information Fusion for Multiple Motion Computation

TO WHOM IT MAY CONCERN, THE FOLLOWING IS  
A SPECIFICATION OF THE AFORESAID INVENTION

DENSITY ESTIMATION-BASED INFORMATION FUSION FOR MULTIPLE MOTION  
COMPUTATION

CROSS-REFERENCE TO RELATED APPLICATIONS

5

The application claims priority to U.S. Provisional Application No. 60/416,835, which was filed on October 8, 2002.

**BACKGROUND OF THE INVENTION**

10 1. Field of the Invention:

The present invention relates to information fusion, and more particularly to nonparametric information fusion for motion estimation.

2. Discussion of Related Art:

15 Information fusion is important for many computer vision tasks. Information fusion is also important across modalities, for applications such as collision warning and avoidance, and speaker localization. Typically, a classical estimation framework such as the extended Kalman filter is employed to  
20 derive an estimate from multiple sensor data.

The problem of information fusion appears in many forms in computer vision. Tasks such as motion estimation, multimodal registration, tracking, and robot localization, typically use a synergy of estimates coming from multiple sources. However,  
25 typically the fusion algorithms assume a single source model and

are not robust to outliers. If the data to be fused follow different underlying models, the conventional algorithms would produce poor estimates.

The quality of information fusion depends on the uncertainty of cross-correlation data. Let  $\hat{x}_1$  and  $\hat{x}_2$  be two estimates that are to be fused together to yield an optimal estimate  $\hat{x}$ . The error covariances are defined by

$$\mathbf{P}_{ij} = E[(\mathbf{x} - \hat{\mathbf{x}}_i)(\mathbf{x} - \hat{\mathbf{x}}_j)^T] \quad (1)$$

for  $i=1,2$  and  $j=1,2$ . To simplify the notation denote  $\mathbf{P}_{11} \equiv \mathbf{P}_1$  and  $\mathbf{P}_{22} \equiv \mathbf{P}_2$ .

Ignoring the cross-correlation,  $\mathbf{P}_{12} \equiv \mathbf{P}_{21}^T = 0$ , the best linear unbiased estimator (BLUE), also called Simple Convex Combination, is expressed by:

$$\hat{\mathbf{x}}_{CC} = \mathbf{P}_{CC} (\mathbf{P}_1^{-1} \hat{\mathbf{x}}_1 + \mathbf{P}_2^{-1} \hat{\mathbf{x}}_2) \quad (2)$$

$$\mathbf{P}_{CC} = (\mathbf{P}_1^{-1} + \mathbf{P}_2^{-1})^{-1} \quad (3)$$

When the initial estimates are correlated  $\mathbf{P}_{12} \equiv \mathbf{P}_{21}^T \neq 0$  and the noise correlation can be measured, the BLUE estimator  $(\hat{x}_{BC}, \mathbf{P}_{BC})$  is derived according to Bar-Shalom and Campo using a Kalman formulation. The most general case of BLUE estimation also assumes prior knowledge of the covariance of  $\mathbf{x}$ .

A conservative approach to information fusion has been proposed by Julier and Uhlman in the form of the Covariance Intersection algorithm. The objective of the Covariance Intersection algorithm was to obtain a consistent estimator of the covariance matrix when two random variables are linearly combined and their cross-correlation is unknown. Consistency means that the estimated covariance is always an upper-bound, in the positive definite sense, of the true covariance, no matter what the cross-correlation level is. The intersection is characterized by the convex combination of the covariances

$$\hat{\mathbf{x}}_{CI} = \mathbf{P}_{CI} (\omega \mathbf{P}_1^{-1} \hat{\mathbf{x}}_1 + (1 - \omega) \mathbf{P}_2^{-1} \hat{\mathbf{x}}_2) \quad (4)$$

$$\mathbf{P}_{CI} = (\omega \mathbf{P}_1^{-1} + (1 - \omega) \mathbf{P}_2^{-1})^{-1} \quad (5)$$

where  $\omega \in [0, 1]$ . The parameter  $\omega$  is chosen to optimize the trace or determinant of  $\mathbf{P}_{CI}$ .

Covariance Intersection has a very suggestive geometrical interpretation: if one plots the covariance ellipses  $\mathbf{P}_1$ ,  $\mathbf{P}_2$  and  $\mathbf{P}_{BC}$  (as given by the Bar-Shalom/Campo formulation) for all choices of  $\mathbf{P}_{12}$ , then  $\mathbf{P}_{BC}$  always lies within the intersection of  $\mathbf{P}_1$  and  $\mathbf{P}_2$ . Thus, the strategy determines a  $\mathbf{P}_{CI}$  that encloses the intersection region and is consistent even for unknown  $\mathbf{P}_{12}$ . It has been shown in that the difference between  $\mathbf{P}_{CI}$  and the true covariance of  $\mathbf{x}$  is a semipositive matrix. More recently, Chong and Mori examined the performance of Covariance Intersection,

while Chen, Arambel and Mehra analyzed the optimality of the algorithm.

Observe that the Covariance Intersection can be generalized to the fusion of  $n$  estimates as

$$\hat{\mathbf{x}}_{CI} = \mathbf{P}_{CI} \sum_{i=1}^n \omega_i \mathbf{P}_i^{-1} \hat{\mathbf{x}}_i \quad (6)$$

$$\mathbf{P}_{CI} = \left( \sum_{i=1}^n \omega_i \mathbf{P}_i^{-1} \right)^{-1} \quad (7)$$

with  $\sum_{i=1}^n \omega_i = 1$ . In equations 6 and 7 the weights  $\omega_i$  are also chosen to minimize the trace or determinant of  $\mathbf{P}_{CI}$ .

Although important from theoretical viewpoint, Covariance Intersection has at least two weaknesses: it assumes a single source model and is not robust to outliers.

Therefore, a need exists for a system and method for information fusion that accommodates multiple source models and is robust to outliers.

#### SUMMARY OF THE INVENTION

According to an embodiment of the present invention, an information fusion method comprises determining a plurality of initial estimates comprising mean and covariance information, the plurality of initial estimates corresponding to a data, and

determining a density based fusion estimate as a most significant mode of a density function determined from the plurality of initial estimates.

The density function is determined according to a variable-  
5 bandwidth kernel density estimate. The method comprises  
determining a covariance of the fusion estimate, wherein the  
covariance is a convex combination of covariances of the  
plurality of initial estimates, each initial estimate having a  
weight, wherein the weights of the initial estimates are  
10 determined according to the most significant mode of the data.  
The most significant mode is determined by mode tracking across  
scales, the mode tracking beginning with a unique mode defined at  
a relatively large scale and tracking that mode toward a  
relatively smaller scale.

15 Determining the plurality of initial estimates comprises  
determining a window bounding the data for motion estimation.  
The method further comprises determining a matrix of spatial  
gradients in the data, determining a vector of temporal image  
gradients in the data, and determining the covariance information  
20 as a matrix proportional to a variance of noise in the data based  
on the matrix of spatial gradients and the vector of temporal  
image gradients.

According to an embodiment of the present invention, a  
method for nonparametric information fusion for motion estimation

comprises determining a plurality of initial estimates of motion,  
tracking a mode of a density function across scales given  
multiple source models, wherein each source model corresponds to  
a set of initial estimates of motion, and determining a location  
5 of a most significant mode of the density function from a fusion  
of the initial estimates, wherein the most significant mode is a  
motion estimate.

The method comprises determining a covariance of the fusion  
of the initial estimates, wherein the covariance is a convex  
10 combination of covariances of the plurality of initial estimates,  
each initial estimate having a weight, wherein the weights of the  
initial estimates are determined according to the most  
significant mode of the data. The most significant mode is  
determined by mode tracking across scales starting from a unique  
15 mode defined at a large scale and tracking that mode toward  
smaller scales.

The density function is determined as a sum value of a  
plurality of Gaussian kernels located at data points in a  
predetermined neighborhood of data. Each Gaussian kernel  
20 comprises spread information and orientation information.

According to an embodiment of the present invention, a  
program storage device is provided, readable by machine, tangibly  
embodying a program of instructions executable by the machine to  
perform method steps for information fusion. The method

comprising determining a plurality of initial estimates  
comprising mean and covariance information, the plurality of  
initial estimates corresponding to a data, and determining a  
density based fusion estimate as a most significant mode of a  
5 density function determined from the plurality of initial  
estimates.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of the present invention will be  
10 described below in more detail, with reference to the  
accompanying drawings:

Figure 1 is a diagram of a system according to an embodiment  
of the present invention;

Figure 2a is a graph of input data represented as ellipses  
15 with 95% confidence according to an embodiment of the present  
invention;

Figure 2b is a graph of fusion results overlaid on input  
data according to an embodiment of the present invention;

Figures 2c-f are graphs showing density surfaces  
20 corresponding to equation (8) with different values for •  
according to an embodiment of the present invention;

Figure 3a shows frame 9 of the New-Sinusoid1 sequence;

Figure 3b shows a correct flow of Figure 3a according to an  
embodiment of the present invention;

25 Figure 3c shows a VBDF flow according to an embodiment of  
the present invention;

Figure 3d shows error corresponding to Figures 3b and 3C, according to an embodiment of the present invention;

Figure 4 is a graph of ellipses with 95% confidence representing initial flow estimates for the location (49,13) of the top level of Yosemite pyramid according to an embodiment of the present invention; and

Figure 5 is a flow chart of a method according to an embodiment of the present invention.

#### DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

According to an embodiment of the present invention, a nonparametric approach to information fusion is called Variable-Bandwidth Density-based Fusion (VBDF). A fusion estimator is determined as the location of the most significant mode of a density function, which takes into account the uncertainty of estimates to be fused. According to an embodiment of the present invention, a method utilizes a variable-bandwidth mean shift determined at multiple scales. The fusion estimator is consistent and conservative, while handling naturally outliers in the data and multiple source models. Experimental results for a fusion estimator according to an embodiment of the present invention are shown for the task of multiple motion estimation, however, it should be noted that a fusion estimator can be applied to other fields, such as collision avoidance, multi-modal

registration of medical data, tracking, robotic localization, and heart motion estimation.

For purposes of the description, multiple sensors provide sensor measurements. Each sensor measurement is characterized by its mean vector and a covariance matrix defining an uncertainty of the mean. When the processing of all measurements takes place at a single location, the fusion is called centralized. In centralized fusion the sensor measurement errors are typically considered independent across sensors and time. A construction with improved reliability and flexibility is provided by distributed fusion, represented by a collection of processing nodes that communicate with each other. Such architecture handles the information as follows: the sensor measurements are evaluated and the state information from a local neighborhood is fused. An important topic in distributed fusion is the handling of cross-correlation, which is difficult to evaluate. The Covariance Intersection algorithm provides a consistent and conservative solution to this problem.

The distributed fusion architecture is suitable for the task of motion estimation from image sequences. According to an embodiment of the present invention, it is assumed that some image property, such as the brightness, is conserved locally in time constrains the component of the motion field in the direction of the spatial image gradient. The initial motion

estimates from a given neighborhood or portion of data are fused to exploit spatial coherence.

It is to be understood that the present invention may be implemented in various forms of hardware, software, firmware, special purpose processors, or a combination thereof. In one embodiment, the present invention may be implemented in software as an application program tangibly embodied on a program storage device. The application program may be uploaded to, and executed by, a machine comprising any suitable architecture.

Referring to Fig. 1, according to an embodiment of the present invention, a computer system 101 for implementing the present invention can comprise, *inter alia*, a central processing unit (CPU) 102, a memory 103 and an input/output (I/O) interface 104. The computer system 101 is generally coupled through the I/O interface 104 to a display 105 and various input devices 106 such as a mouse and keyboard. The support circuits can include circuits such as cache, power supplies, clock circuits, and a communications bus. The memory 103 can include random access memory (RAM), read only memory (ROM), disk drive, tape drive, etc., or a combination thereof. The present invention can be implemented as a routine 107 that is stored in memory 103 and executed by the CPU 102 to process the signal from the signal source 108. As such, the computer system 101 is a general

purpose computer system that becomes a specific purpose computer system when executing the routine 107 of the present invention.

The computer platform 101 also includes an operating system and micro instruction code. The various processes and functions  
5 described herein may either be part of the micro instruction code or part of the application program (or a combination thereof) which is executed via the operating system. In addition, various other peripheral devices may be connected to the computer platform such as an additional data storage device and a printing  
10 device.

It is to be further understood that, because some of the constituent system components and method steps depicted in the accompanying figures may be implemented in software, the actual connections between the system components (or the process steps)  
15 may differ depending upon the manner in which the present invention is programmed. Given the teachings of the present invention provided herein, one of ordinary skill in the related art will be able to contemplate these and similar implementations or configurations of the present invention.

20 An adaptive density estimation with variable kernel bandwidth can be applied in computer vision. Variable-bandwidth methods improve the performance of kernel estimators by adapting the kernel scaling and orientation to the local data statistics.

Let  $x_i, i=1 \dots n$ , be  $n$  data points in the  $d$ -dimensional space  $R^d$ . By selecting a different bandwidth matrix  $\mathbf{H}_i = \mathbf{H}(\mathbf{x}_i)$  (e.g., assumed full rank) for each  $x_i$  we define the sample point density estimator is defined as:

$$\hat{f}_v(\mathbf{x}) = \frac{1}{n(2\pi)^{d/2}} \sum_{i=1}^n \frac{1}{|\mathbf{H}_i|^{1/2}} \exp\left(-\frac{1}{2} D^2(\mathbf{x}, \mathbf{x}_i, \mathbf{H}_i)\right) \quad (8)$$

where

$$D^2(\mathbf{x}, \mathbf{x}_i, \mathbf{H}_i) \equiv (\mathbf{x} - \mathbf{x}_i)^\top \mathbf{H}_i^{-1} (\mathbf{x} - \mathbf{x}_i) \quad (9)$$

is the Mahalanobis distance from  $\mathbf{x}$  to  $\mathbf{x}_i$ . The variable-bandwidth mean shift vector at location  $\mathbf{x}$  is given

by:

$$\mathbf{m}_v(\mathbf{x}) \equiv \mathbf{H}_h(\mathbf{x}) \sum_{i=1}^n \omega_i(\mathbf{x}) \mathbf{H}_i^{-1} \mathbf{x}_i - \mathbf{x} \quad (10)$$

where  $\mathbf{H}_h$  is the data-weighted harmonic mean of the bandwidth matrices determined at  $\mathbf{x}$

$$\mathbf{H}_h(\mathbf{x}) = \left( \sum_{i=1}^n \omega_i(\mathbf{x}) \mathbf{H}_i^{-1} \right)^{-1} \quad (11)$$

and

$$\omega_i(\mathbf{x}) = \frac{\frac{1}{|\mathbf{H}_i|^{1/2}} \exp\left(-\frac{1}{2} D^2(\mathbf{x}, \mathbf{x}_i, \mathbf{H}_i)\right)}{\sum_{i=1}^n \frac{1}{|\mathbf{H}_i|^{1/2}} \exp\left(-\frac{1}{2} D^2(\mathbf{x}, \mathbf{x}_i, \mathbf{H}_i)\right)} \quad (12)$$

are weights satisfying  $\sum_{i=1}^n \omega_i(\mathbf{x}) = 1$ . It can be shown that the iterative computation of the mean shift vector (10) always moves

the point  $x$  to a location where the density (8) is higher or equal to the density at a previous location. As a result, an iterative hill-climbing procedure is defined, which converges to a stationary point, e.g., zero gradient, of the underlying  
 5 density.

The VBDF estimator is defined as the location of the most significant sample mode of the data.

For determination through multiscale optimization, assume that the data points  $x_i, i=1\dots n$  are each associated with a  
 10 covariance matrix  $C_i$  that quantifies uncertainty. The location of the most significant mode is determined in a multiscale fashion, by tracking the mode of the density function across scales.

More specifically a first mode detection is performed using  
 15 large bandwidth matrices of the form  $H_i=C_i+\alpha^2I$ , where the parameter  $\alpha$  is large with respect to the spread of the points  $x_i$ . The mode detection method is based on mean shift and involves the iterative computation of expression (10) and translation of  $x$  by  $m_v(x)$  until convergence. At the largest scale, the mode location  
 20 does not depend on the initialization, up to some numerical approximation error, since for large  $\alpha$  the density surface is unimodal. In the next stages, the detected mode is tracked across scales by successively reducing the parameter  $\alpha$  and

performing mode detection again. At each scale the mode detection algorithm is initialized with the convergence location from the previous scale.

Note that for the last mode detection procedure, the bandwidth matrix associated with each data point is equal to the point covariance matrix, where,  $\mathbf{H}_i = \mathbf{C}_i$ ,  $i = 1 \dots n$ . Denote by  $\hat{\mathbf{x}}_m$  the location of the most significant mode. Since the gradient at  $\hat{\mathbf{x}}_m$  is zero,  $m_v(\hat{\mathbf{x}}_m) = 0$ , which means

$$\hat{\mathbf{x}}_m = \mathbf{H}_h(\hat{\mathbf{x}}_m) \sum_{i=1}^n \omega_i(\hat{\mathbf{x}}_m) \mathbf{H}_i^{-1} \mathbf{x}_i \quad (13)$$

$$\mathbf{H}_h(\hat{\mathbf{x}}_m) = \left( \sum_{i=1}^n \omega_i(\hat{\mathbf{x}}_m) \mathbf{H}_i^{-1} \right)^{-1} \quad (14)$$

A VBDF estimator can be expressed as Equations (13) and (14), wherein the VBDF estimator has the following properties: the covariance (14) of the fusion estimate is a convex combination of the covariances of initial estimates. The matrix

$\mathbf{H}_h(\hat{\mathbf{x}}_m)$  is a consistent and conservative

estimate of the true covariance matrix of  $\hat{\mathbf{x}}_m$ , irrespective of the actual correlations between initial estimates. The criterion of a method according to an embodiment of the present invention is based on the most probable value of the data, e.g., the most significant mode. This criterion can be used when the data is

multimodal, where the initial estimate belong to different source models. Such a property is common for motion estimation since the points in a local neighborhood may exhibit multiple motions. The most significant mode corresponds to the most relevant motion. The tracking of the density mode across scales insures the detection of the most significant mode. The use of a Gaussian kernel is used for the continuity of the modes across scales. Also, by selecting the most significant mode, the estimate is also robust to outliers.

Comparing, experimentally, the new VBDF estimator against the BLUE and Covariance Intersection, a synthetic input data is shown in Figure 2a and comprises eight initial bi-variate estimates expressed as location and covariance. Each covariance is displayed as an ellipse with 95% confidence. Trajectory of mode tracking across scales is shown 201. An ellipse corresponding to a VBDF 202 estimate is drawn with a thick line. Observe that the input data has a clearly identifiable structure of five measurements, while the other three measurements can be considered outliers. In addition, the uncertainty of the data is low and the mean vectors are far apart from each other. This creates a difficult mode estimation problem.

The same figure shows the VBDF estimate, having a mean equal to  $(-0.3499, 0.1949)$  and its covariance, represented by an ellipse of thick line. The VBDF ellipse masks one of the input ellipses.

The trajectory of the mode tracking across scales is also plotted. Each small circle indicates the result of mode detection for one scale.

In Figure 2b a result of the VBDF is compared with that of the BLUE fusion ((2) and (3)) and Covariance Intersection ((6) and (7)). In Figure 2b, fusion results are overlaid on input data. Ellipses are represented with squares for BLUE estimate 203 and diamonds for Covariance Intersection 204. The kernel density estimate determined with  $H_i = C_i + \alpha^2 I$  is shown in Figures 2c-10 f for different values of  $\alpha$ . A triangle marks the location of the most significant mode across scales. Figure 2f is obtained with  $H_i = C_i$  and corresponds to a VBDF estimate.

The following conclusions can be drawn: the BLUE estimate produces the most confident result, however, the presence of outliers in the data has a strong, negative influence on this estimate. At the same time the BLUE estimate can be overly confident by neglecting the cross-correlation. The Covariance Intersection is also negatively influenced by outliers. The weights have been optimized to minimize the trace of the covariance matrix. However, since the optimization regards only 20 the covariance and not the location, the resulting estimate is rather poor. Note that by employing the variable-bandwidth mean shift and mode tracking across scales, the VBDF uses an optimization of the weights. Observe that, as expected, the VBDF

method has not been influenced by outliers. Inferring from Figure 2c, the most significant mode across scales is not the highest mode determined with the bandwidths  $H_i=C_i$ ! Note the highest location on the density landscape determined with  $H_i=C_i$  is located at  $(0.2380, -1.333)$ , which is different from the VBDF estimate. This conclusion is in agreement with an expectation that the most significant mode should not be determined based solely on local information. The multiscale method makes the right choice in selecting the right mode.

For the estimation of multiple motion, an application of the VBDF estimator is adapted to the computation of multiple motion. Detailed reviews on motion estimation are given by Aggarwal and Nandhakumar, Mitiche and Bouthemy, and Nagel. Three main approaches to motion estimation can be identified, based on spatial gradient, image correlation, and regularization of spatio-temporal energy. The motion is commonly assumed to be locally constant, affine, or quadratic.

Many of the techniques based on spatial gradient embrace a two step approach for the computation of motion flow. An initial estimate of the flow is determined for each image location using the brightness constancy. The initial estimates are then fused locally in the hope for a better fusion estimate. The presence of multiple motions, however, makes the second task difficult since the initial estimates are generated by multiple and unknown

source models. Multiple motions can be generated by objects moving with different velocities, but can also be the result of transparency, highlights or shadows.

One of the most popular and efficient optical flow techniques has been developed by Lucas and Kanade in the context of stereo vision. They neglected the uncertainty of initial estimates and use (weighted) least squares in a neighborhood to fuse them. Weber and Malik employed the total least squares for the same task. Simoncelli, Adelson and Heeger improved the method by determining and using the uncertainty of initial estimates. Nevertheless, they assume that the initial estimates are independent and do not model multiple motions. Black and Anandan approached the motion estimation problem in a robust framework, being able to deal with multiple motions.

The first benchmarking effort on the evaluation of motion estimation algorithms has been conducted by Barron, Fleet and Beuchemin. Since then, many of the newly proposed methods are compared using the Barron, Fleet and Beuchemin methodology, as is presented herein with respect to the experimental data achieved for a VBDF method according to an embodiment of the present invention.

For a given image location we extract an initial motion estimate is extracted from a very small  $N \times N$  neighborhood using Biased Least Squares (BLS)

$$\hat{\mathbf{x}} = (\mathbf{A}^\top \mathbf{A} + \beta \mathbf{I})^{-1} \mathbf{A}^\top \mathbf{b} \quad (15)$$

where  $\mathbf{A}$  is the  $N^2 \times 2$  matrix of spatial image gradients, and  $\mathbf{b}$  is the  $N^2$ -dimensional vector of temporal image gradients.

The BLS solution has a covariance matrix  $\mathbf{C}$  that is  
 5 proportional to the variance  $\sigma^2$  of the image noise. The  
 advantage of BLS is that it avoids instability problems in the  
 regular Least Squares solution by allowing a small amount of  
 bias. The technique is also called ridge regression or Tikhonov  
 regularization and various solutions have been proposed to  
 10 compute the regularization parameter  $\beta$  from the data.

The motion flow information is combined in a local image  
 neighborhood of dimension  $n=M \times M$  using the VBDF estimator ((13)  
 and (14)). Denoting by  $(\hat{x}_i, C_i), i=1 \dots n$  the initial flow estimates  
 produced through BLS, their fusion results in

$$\hat{\mathbf{x}}_m = \mathbf{C}(\hat{\mathbf{x}}_m) \sum_{i=1}^n \omega_i(\hat{\mathbf{x}}_m) \mathbf{C}_i^{-1} \hat{\mathbf{x}}_i \quad (16)$$

$$\mathbf{C}(\hat{\mathbf{x}}_m) = \left( \sum_{i=1}^n \omega_i(\hat{\mathbf{x}}_m) \mathbf{C}_i^{-1} \right)^{-1} \quad (17)$$

where

$$\omega_i(\hat{\mathbf{x}}_m) = \frac{\frac{1}{|\mathbf{C}_i|^{1/2}} \exp\left(-\frac{1}{2} D^2(\hat{\mathbf{x}}_m, \hat{\mathbf{x}}_i, \mathbf{C}_i)\right)}{\sum_{i=1}^n \frac{1}{|\mathbf{C}_i|^{1/2}} \exp\left(-\frac{1}{2} D^2(\hat{\mathbf{x}}_m, \hat{\mathbf{x}}_i, \mathbf{C}_i)\right)} \quad (18)$$

and  $\hat{\lambda}_m$  is determined through mode tracking across scales, as discussed above as the determination through multiscale optimization.

Regarding the experiments, a three level image pyramid was  
 5 constructed using a five-tap filter [0.0625\ 0.25\ 0.375\ 0.25\ 0.0625]. For the derivative filters in both spatial and temporal domain we used the simple difference. As a result, the optical flow was determined from three frames, from coarse to fine. Initial flow estimates were obtained in a neighborhood of three  
 10 (i.e.,  $N=3$ ) and the regularization parameter  $\beta=1$ . Estimation errors were evaluated using the software that determines the average angular error  $\mu_e$  and its standard deviation  $\sigma_e$ . Here, the flow estimated with a density of 100 is discussed.

A first test involved the sequence New-Sinusoid1 introduced  
 15 by Bab-Hadiashar and Suter. This sequence (see Figure 3a) has spatial frequencies similar to Sinusoid1 from but has a central stationary square of 50 pixels, thus containing motion discontinuities. The correct flow for New-Sinusoid1 is shown in Figure 3b. The robust motion estimation method described in has  
 20 errors in the range  $(\mu_e=1.51-2.82, \sigma_e=5.86-8.82)$ .

Using VBDF estimation according to an embodiment of the present invention, a substantially decrease in errors was obtained to  $(\mu_e=0.57, \sigma_e=5.2)$  and the estimated motion has sharp

boundaries (Figure 3c). In Figure 3d the angular error is shown multiplied by 100 (white space corresponds to large errors).

These results were obtained with a seven analysis window (e.g.,  $M=7$ ) and variable bandwidth mean shift applied across five

5 scales. The noise variance used in BLS has been assumed to be  $\sigma^2=0.08$ , equal to that of the quantization noise. An average number of three mean shift iterations per scale per window were executed.

A second test was performed using the Yosemite sequence.

10 This synthetic sequence comprises many challenges, including multiple motions and aliasing. Numerous results have been reported on Yosemite involving either the complete sequence, or the partial sequence, with the sky and clouds discarded. For the complete sequence, a VBDF method according to an embodiment of  
15 the present invention resulted in  $(\mu_e = 4.25, \sigma_e = 7.82)$  for the middle frame. The estimated flow is shown in 3a and 3b presents the angular error.

In Figure 4 a fusion example is shown for Yosemite corresponding to the location (49,13) at the top of the image  
20 pyramid. A window of  $M=5$  has been used to collect 25 initial estimates. The starting point of a method according to an embodiment of the present invention is represented by a large dot in the center. This location is situated at the border between

the sky and mountain. The initial location of the mode detection algorithm is marked by a large dot. The VBDF ellipse, according to an embodiment of the present invention, is drawn with a thick line.

5 For the skyless Yosemite, a VBDF method according to an embodiment of the present invention obtained  $(\mu_e = 1.55, \sigma_e = 1.65)$ .

In comparison to other techniques, a VBDF method is simple, easy to implement, and efficient, being based on the detection of the most significant mode of the density of some initial  
 10 estimates. For Yosemite, a 15x15 analysis window and  $\sigma^2 = 7.82$  were used. In addition, the distances between the initial flow vectors were weighted according to the intensity difference between the corresponding image pixels by a Gaussian kernel of standard deviation equal to twelve. This assured that flow  
 15 vectors similar in direction and magnitude and coming from locations with similar intensity were grouped together.

For the Translating Tree, a VBDF method according to an embodiment of the present invention obtained  $(\mu_e = 0.19, \sigma_e = 0.17)$ .  
 For the Diverging Tree a VBDF method according to an embodiment  
 20 of the present invention resulted in  $(\mu_e = 1.10, \sigma_e = 0.73)$ .

Resulting flow for the SRI sequence is presented in Figure 7. Observe the sharp flow boundaries. The same parameters as in

Yosemite were used for these sequences, but without intensity weighting.

According to an embodiment of the present invention, a VBDF estimator is provided as a powerful tool for information fusion  
5 based on adaptive density estimation. A fusion estimator can handle with multiple source models and to handle cross-correlation in a consistent way. Comparing the VBDF framework with the BLUE fusion and Covariance Intersection showed that the new estimator can be used to construct a very effective motion  
10 computation method.

The detection of the most significant mode of a density function is accomplished through mode tracking across scales, that is each successive estimation uses a previous estimation as a starting point. Referring to Figure 5, a method for  
15 information fusion comprises determining a plurality of initial estimates of motion 501, tracking a mode of a density function across scales given multiple source models, wherein each source model corresponds to an initial estimate of motion 502, and determining a location of a most significant mode of the density  
20 function from a fusion of the initial estimates. In the context of motion estimation, the most significant mode corresponds to the most relevant motion in the local neighborhood 503. The same concepts can be naturally extended to other vision domains such as stereo, tracking, or robot localization.

Having described embodiments for nonparametric information fusion for motion estimation, it is noted that modifications and variations can be made by persons skilled in the art in light of the above teachings. It is therefore to be understood that

5 changes may be made in the particular embodiments of the invention disclosed which are within the scope and spirit of the invention as defined by the appended claims. Having thus described the invention with the details and particularity required by the patent laws, what is claimed and desired

10 protected by Letters Patent is set forth in the appended claims.